SCIENCE CHINA
Information Sciences

• LETTER •

January 2023, Vol. 66 119104:1–119104:2
https://doi.org/10.1007/s11432-022-3541-x

BiTGAN: bilateral generative adversarial networks for Chinese ink wash painting style transfer

Xiao HE[1], Mingrui ZHU[1*], Nannan WANG[1*], Xiaoyu WANG[2] & Xinbo GAO[3]

[1]*State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an* 710071*, China;*
[2]*The Chinese University of Hong Kong (Shenzhen), Shenzhen* 518172*, China;*
[3]*Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing* 400065*, China*

Received 9 February 2022/Revised 25 May 2022/Accepted 27 June 2022/Published online 1 November 2022

Citation He X, Zhu M R, Wang N N, et al. BiTGAN: bilateral generative adversarial networks for Chinese ink wash painting style transfer. Sci China Inf Sci, 2023, 66(1): 119104, https://doi.org/10.1007/s11432-022-3541-x

Dear editor,

Chinese ink wash painting occupies a pivotal position in Chinese traditional painting and has high artistic value. It is meaningful to design a special automatic algorithm for Chinese ink wash painting style transfer. In existing algorithms, there are two types of methods that can achieve this goal. One is the image style transfer method, which can achieve instance-level style transfer according to the different reference images. However, because of the significant difference between Chinese ink wash painting and Western painting techniques, directly applying the existing image style transfer methods to the style transfer of Chinese ink wash painting will produce undesired results. The other is the image-to-image translation method, which can realize domain-level style transfer. For the style transfer of Chinese ink wash painting, ChipGAN [1] summarizes the unique painting skills of Chinese ink wash painting and designs corresponding loss for transfer. Despite the demonstrated improvement, it still has some limitations: (1) the expression form of voids is unreasonable and lacks context information; (2) it fails to keep a reasonable balance between content and style, which destroys the original structure of the image in the process of style transfer. To address these problems, we combine the complementary strength of the image style transfer method and image-to-image translation method and design a novel architecture dubbed bilateral generative adversarial networks (BiTGAN) to achieve better performance in the style transfer of Chinese ink wash painting.

BiTGAN. Based on CycleGAN [2], we present a new bilateral generator for Chinese ink wash painting style transfer. The architecture of the bilateral generator is illustrated in Figure 1. It consists of four modules: the weight-shared encoder, the UNet Path [3], the ResNet Path [4], and the decoder.

The design intention of the weight-shared encoder is clear. It can not only reduce the number of parameters but also provide a feature-shared distribution for ResNet Path and UNet Path by using shared convolutional layers. Specifically, the weight-shared encoder is used to encode the source image $I_{\rm s}$. It consists of three weight-shared convolutional layers, which encode the input image to the shared feature $F_{\rm ws}$:

$$F_{\rm ws} = {\rm Enc}\left(I_{\rm s}\right), \tag{1}$$

where Enc denotes the weight-shared encoder, $F_{\rm ws} \in \mathfrak{R}^{(C\times H\times W)}$, C, H, W denote the channel number, height and width of the feature maps.

UNet Path mainly consists of a contracting path and a symmetric expanding path. The contracting path helps the generator capture more contextual information. And the symmetric extending path can combine the features of the contracting path with the features of up-sampled output to obtain a multi-scale feature. In our task, the contracting path can preserve reasonable voids in the output image, and the symmetric expanding path can preserve the global structure (e.g., pose) of the source image. However, because deep- and shallow-layer feature maps with low and high resolutions in the decoder indicate different levels of semantic information from high-level structures to low-level colors/textures [5], the shallow-symmetric expanding path may transfer low-level information (e.g., color) from the source image to the output image. To this end, our UNet Path is designed for middle and deep layers. We take the shared feature $F_{\rm ws}$ as the input and generate the middle feature $F_{\rm u}$ rather than the output image. Under the loss constraint, $F_{\rm u}$ will contain the structural information and context information.

$$F_{\rm u} = {\rm UP}\left(F_{\rm ws}\right), \tag{2}$$

where UP denotes the UNet Path and $F_{\rm ws}$ is the shared feature.

Recent image translation methods [4] transfer the style of the target domain by equipping ResBlocks in the bottleneck of the generator. Similarly, we equip our RP with several ResBlocks after the weight-shared encoder. Through a series of ResBlocks, we get the feature $F_{\rm r}$ that derives from

* Corresponding author (email: mrzhu@xidian.edu.cn, nnwang@xidian.edu.cn)

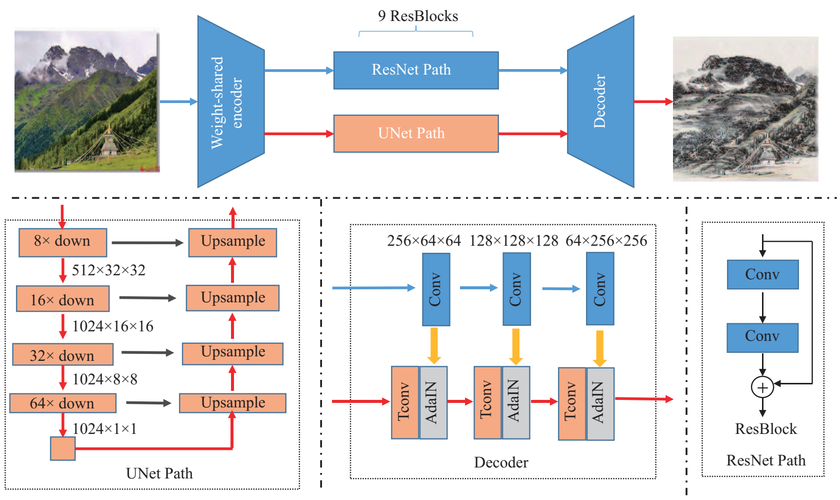

Figure 1 (Color online) The pipeline of the proposed BiTGAN for Chinese ink wash painting style transfer. Given an input image, the weight-shared encoder is utilized to encode the image and obtain the features. Then the UNet Path down-samples the feature maps to obtain a large receptive field, which encodes high level semantic context information. The ResNet Path uses a series of residual blocks (ResBlocks) to extract the low-level information (e.g., color, texture). In the decoder, the ResNet Path normalizes the features derived from UNet Path according to the AdaIN layer, and then the normalized feature is up-sampled to generate Chinese ink wash painting results.

RP.

$$F_{\mathrm{r}} = \mathrm{RP}\left(F_{\mathrm{ws}}\right), \tag{3}$$

where RP denotes the ResNet Path.

In the decoding process, we take the features F_{u} and F_{r} as the input of adaptive instance normalization (AdaIN) [6], which aligns the channel-wise mean and variance of F_{u} to match those of F_{r}. Let F_{n} denote this normalization feature. Then we up-sample the features F_{r} and F_{n} and perform adaptive instance normalization in the same time so that F_{n} can obtain global style patterns from F_{r}. Finally, we reconstruct the output image from the latest normalization feature F'_{n}. F'_{n} has the same mean and variance as F_{r}, so when the adversarial loss pushes the output image to conform to the distribution of the Chinese ink wash painting domain, RP will obtain the global style patterns from the Chinese ink wash painting domain. It means that through the adaptive instance normalization layer, RP can help UP supplement the style patterns to achieve visually pleasing Chinese ink wash painting style transfer.

$$F_{\mathrm{n}} = \mathrm{AdaIN}\left(F_{\mathrm{u}}, F_{\mathrm{r}}\right), \tag{4}$$

$$\mathrm{AdaIN}\left(F_{\mathrm{u}}, F_{\mathrm{r}}\right) = \sigma\left(F_{\mathrm{r}}\right)\left(\frac{F_{\mathrm{u}} - \mu\left(F_{\mathrm{u}}\right)}{\sigma\left(F_{\mathrm{u}}\right)}\right) + \mu\left(F_{\mathrm{r}}\right), \tag{5}$$

$$I_{\mathrm{o}} = \mathrm{Dec}\left(F'_{\mathrm{n}}\right), \tag{6}$$

where σ and μ are the channel-wise mean and standard deviation; Dec denotes the decoder.

Experiments. The overall model consists of the proposed bilateral generator and a discriminator. For discriminator networks, we use 70 × 70 PatchGAN [7]. The BiTGAN was trained with adversarial loss and cycle consistency loss on the "ChipPhi" dataset [1]. We trained BiTGAN for 200 epochs using Adam optimizer with a batch size of 1. Detailed experimental results and analysis can be found in Appendixes A–C.

Conclusion. In this study, we propose a novel BiTGAN model for Chinese ink wash painting style transfer. The generator of BiTGAN consists of a UNet Path and a ResNet Path. Under the loss constraint, UNet Path focuses on extracting context information and preserving the content structure. ResNet Path aligns the mean and variance to supplement the global style patterns. Extensive experiments demonstrate that our method achieves superior performance to state-of-the-art methods in the Chinese ink wash painting style transfer.

Acknowledgements This work was supported in part by National Key Research and Development Program of China (Grant No. 2018AAA0103202) and National Natural Science Foundation of China (Grant Nos. 62106184, 62036007, 61922066, 61876142, 62176198).

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 He B, Gao F, Ma D, et al. ChipGAN: a generative adversarial network for Chinese ink wash painting style transfer. In: Proceedings of the 26th ACM International Conference on Multimedia, 2018. 1172–1180
2 Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, 2017. 2223–2232
3 Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer-assisted Intervention, Cham, 2015. 234–241
4 He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. 770–778
5 Zeiler M D, Fergus R. Visualizing and understanding convolutional networks. In: Proceedings of European Conference on Computer Vision, Cham, 2014. 818–833
6 Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision, 2017. 1501–1510
7 Li C, Wand M. Precomputed real-time texture synthesis with Markovian generative adversarial networks. In: Proceedings of European Conference on Computer Vision, Cham, 2016. 702–716